

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

*In re* Application of: )  
Akira ASAKURA *et al.* )  
Serial No.: 09/470,667 ) Examiner: M. Walicka  
Filed: December 22, 1999 ) Art Unit: 1652  
For: **NOVEL ALCOHOL/ALDEHYDE** )  
**DEHYDROGENASES** )

Commissioner for Patents  
Washington, DC 20231

**SECOND DECLARATION OF DR. MASAKO SHINJOH UNDER 37 C.F.R. § 1.132**

Sir:

I, Masako Shinjoh, a citizen and resident of Japan, hereby declare as follows:

1. I am employed by Nippon Roche Research Center of Nippon Roche K.K., Kajiwara 200, Kamakura-shi, Kanagawa-ken 247-8530, Japan (hereafter "NRKK"). I currently hold the position of genetic engineer at NRKK. A copy of my *curriculum vitae* is attached as Exhibit 1.
2. I am a coinventor of U.S. patent application No: 09/470,667 (the '667 application). The '667 application is summarized in more detail in the FIRST DECLARATION OF DR. MASAKO SHINJOH UNDER 37 C.F.R. § 1.132 ("First Declaration") filed concurrently herewith.

3. As described in the First Declaration, after reviewing the Sequence Listing filed with the '667 application, how the nucleotide and amino acid sequences that make up the Sequence Listing were incorporated into the '667 application, and the original nucleotide printouts from the sequencing machine used to read the experimentally derived sequences, I have come to the conclusion that SEQ ID NOs:1, 3, and 7 each contain a single base (SEQ ID NOs:1 and 3) or a single amino acid (SEQ ID NO:7) error that arose through typing errors.
4. After the '667 application was filed, I found discrepancies in the nucleotide and amino acid sequences identified in the '667 application as SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 7, respectively when compared to the computer printouts generated by the nucleotide sequencing machine used to read the nucleotide sequences that ultimately became SEQ ID NOs:1 and 3 in the '667 application. As set forth in more detail below, I believe that each of these discrepancies was the result of a typing mistake made when I prepared the sequence listing data for an internal Research Report.
5. The original sequence data underlying each of the sequences disclosed in the Sequence Listing of the '667 application were generated by a nucleotide sequencing machine, and could not be converted into an electronic file for manipulation in an electronic

medium (e.g., a word processor). Accordingly, I manually typed the sequences ultimately disclosed in the '667 application into an electronic format using the original sequence data generated by the nucleotide sequencing machine. It is my belief that when the original sequence data was retyped into an electronic format that a single base in each of SEQ ID NOs:1 and 3 was entered in error, and that because of the error in SEQ ID NO:3, its deduced amino acid sequence (SEQ ID NO:7) also contained a single amino acid error. The manually re-typed sequences, including the unrecognized typographical mistakes, were then incorporated into the foreign priority application (EP 96115001 filed September 19, 1996), which became the basis for the '667 application including the Sequence Listing contained therein. (Exhibit 2).

6. A copy of the original printout from the nucleotide sequencing machine of the open reading frame of Enzyme A including the nucleotide sequence (which became SEQ ID NO:1 in the '667 application) and its deduced amino acid sequence (which became SEQ ID NO:5 in the '667 application) is attached as Exhibit 3. I have compared the nucleotide and deduced amino acid sequences from the original printout with the sequences disclosed as SEQ ID NOs:1 and 5 in the '667 application, and have found that the nucleotide at position 852 of SEQ ID NO:1 is a "G" whereas the corresponding nucleotide in the original printout is a "C." It is my

belief that the correct nucleotide at position 852 is "C," not "G" as recited in SEQ ID NO:1.

7. Because of the redundancy of the genetic code, when SEQ ID NO:1 was translated, the deduced amino acid encoded by the codon containing the nucleotide at position 852 did not change compared to the deduced amino acid sequence generated by the nucleotide sequencer as set forth in the original printout. Thus, both sequences are identical.
8. A copy of the original printout from the nucleotide sequencing machine of the open reading frame of Enzyme A" including the nucleotide sequence (which became SEQ ID NO:3 in the '667 application) and its deduced amino acid sequence (which became SEQ ID NO:7 in the '667 application) is attached as Exhibit 4. I have compared the nucleotide and deduced amino acid sequences from the original printout with the sequences disclosed as SEQ ID NOs:3 and 7 in the '667 application and have found that the nucleotide at position 644 of SEQ ID NO:3 is an "A" whereas the corresponding nucleotide in the original printout is a "C." It is my belief that the correct nucleotide at position 644 is "C," not "A" as recited in SEQ ID NO:3.
9. The replacement of "A" for "C" at position 644 in SEQ ID NO:3 also led to the translation of a different amino acid ("Asn" was translated

instead of "Thr" at position 192) from the codon containing the error at nucleotide position 644. It is my belief that the correct amino acid at position 192 of SEQ ID NO:7 is "Thr," not "Asn" as currently recited.

10. To verify the correctness of the nucleotide and amino acid sequences identified on the original printouts generated by the nucleotide sequencing machine, which were the bases for the disclosure of SEQ ID NOs:1, 3, and 7 in the '667 application, I obtained a sample of *Gluconobacter oxydans* strain DSM 4025, the same microorganism from which the nucleotide sequences of SEQ ID NOs:1 and 3 were isolated (as disclosed in the '667 application), from the International Depository Authority, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH ("DSMZ"), a publicly available cell depository.
11. With the assistance of Mr. Naoki Itoh, NRKK's Patent & Licensing Manager, I then contracted with an independent nucleotide sequencing company (Sawady – see the First Declaration) to use the *Gluconobacter oxydans* DSM 4025 cell sample I obtained from DSMZ to clone and sequence the relevant parts of the chromosomal DNA of these cells.
12. The chain of custody of the cell sample and chromosomal DNA derived therefrom is set forth in my First Declaration and the

DECLARATION OF MR. MASAO MASHITA UNDER 37 C.F.R.  
§ 132 and of the DECLARATION OF MR. YOSHITAKA MURATA  
UNDER 37 C.F.R. §1.132, both of which are being filed concurrently  
herewith.

13. With respect to the sequence work, I instructed Sawady to utilize two primer pairs designed by the coinventors for the cloning (by polymerase chain reaction (PCR)) and sequencing of Enzyme A (Primers for Analysis 1) and of Enzyme A" (Primers for Analysis 2) as described below.

Primers for Analysis 1: (for Enzyme A)

Forward: A697f 5' - TACGAAGCCC GTTGGATGAC -3'

Reverse: A1000r 5' - TCGGGTTGAT CGACTGCAGA -3'

Primers for Analysis 2: (for Enzyme A")

Forward: A"479f 5' - TATTCGACGT CGATCGCGGT -3'

Reverse: A"780r 5' - AACTGCTGAG GTGCCGTAGT -3'

14. The Primers for Analysis 1 were designed to amplify (by PCR) the region from nucleotide (nt) position 697 to nt position 1000 of the gene encoding Enzyme A and to determine the amplified nucleotide sequence having 304 bases including the nucleotide at position 852. The primers for Analysis 2 were designed to amplify (by PCR) the region from nt position 479 to nt position 780 of the gene encoding Enzyme A" and to determine the amplified nucleotide sequence having 302 bases including the nucleotide at position 644.

15. The primer information was provided to Mr. Masao Mashita at

Sawady together with a sample of the original microorganism DSM 4025 disclosed in the '667 application (and obtained through DSMZ) to facilitate the cloning and sequencing of the relevant nucleotides for Enzyme A (SEQ ID NO:1) and Enzyme A" (SEQ ID NO:3). (See my First Declaration).

16. On October 13, 2000, I received from Sawady, via Mr. Itoh, an Experimental Report (non-finalized) including the sequence data, which are set forth in Exhibit 5. From the anti-parallel alignment of the (+) and (–) strands in combination with the sequence information of the primers used, I confirmed the correctness of the two nucleotide sequences set forth below. For determining each sequence, I took into consideration that at positions downstream of each primer used in the PCR sequencing carried out by Sawady, the nucleotide reading on the sense strand was not absolutely reliable, and thus for each such region, the data from the complementary sequence was used:

**Ex. (1): A697-1000 Sawady [304 bp] (corresponds to Enzyme A, i.e. SEQ ID NO: 1)**

697

TACGAAGCCC GTTGGATGAC CGGTGCCTGG GGCCAGATCA CCTATGACCC

CGTCACCAAC CTTGTCCACT ACGGCTCGAC CGCTGTGGGT CCGGCGTCGG

AAACCCAACG CGGCACCCCG GGCGGCACGC TGTACGGCAC GAACACCCGT

852

TTCGCCGTGC GTCCTGACAC GGGCGAGATT GTCTGGCGTC ACCAGACCCT

GCCCCGCGAC AACTGGGACC AGGAATGCAC GTTCGAGATG ATGGTCACCA

ATGTGGATGT CCAACCCTCG ACCGAGATGG AAGGTCTGCA GTCGATCAAC



1000  
CCGA

17. The correctness of the above-identified sequence was verified with the two nucleotide sequences (41F903 and 39F903) (Exhibit 5) and two primer sequences (A697f and A1000r):

- \* Nucleotides 697-716 of Ex(1) above are the same as nucleotides 1-20 of primer A697f;
- \* Nucleotides 717-966 of Ex(1) above are the same as nucleotides 41-290 of the complementary sequence of 41F903;
- \* Nucleotides 967-980 of Ex(1) above are the same as nucleotides 253-266 of 39F903; and
- \* Nucleotides 981-1000 of Ex(1) above are the same as nucleotides 1-20 of the complementary sequence of primer A1000r.

**Ex. (2): A"479-780 Sawady [302 bp] (corresponds to Enzyme A", i.e. SEQ ID NO: 3)**

479  
TATTCGACGT CGATCGCGGT CAAGGCACGG ATATGGTCTC GAACTCGTCC  
GGCCCGATTG TCGCCAATGG CGTCATCGTT GCGGGCTCGA CCTGTCAGTA  
TTCGCCGTTC GGCTGTTTCG TTTCGGGCCA CACTCGGCC ACCGGTGAAG  
644  
AGCTGTGGCG CAACACCTTT ATCCCGCGCG CCGGCGAAGA GGGTGATGAG  
ACCTGGGGCA ATGATTACGA GGCCCGCTGG ATGACCGGCG TTTGGGGCCA  
GATCACCTAT GACCCCGTTG GCGGCCTTGT CCACTACGGC ACCTCAGCAG  
780  
TT

18. The correctness of the above-identified sequence was verified with two nucleotide sequences (45F903 and 43F903) (Exhibit 5) and two primer sequences (A"479f and A"780r):

- \* Nucleotides 479-498 of Ex(2) above are the same as nucleotides 1-20 of primer A"479f;
- \* Nucleotides 499-728 of Ex(2) above are the same as nucleotides 41-270 of the complementary sequence of 45F903;
- \* Nucleotides 729-760 of Ex(2) above are the same as nucleotides 228-259 of 43F903; and
- \* Nucleotides 761-780 of Ex(2) above are the same as nucleotides 1-20 of the complementary sequence of primer A"780r.

19. Based on my knowledge and experience, and in view of the results presented herein, it is my opinion that SEQ ID NOs:1 and 3 of the '667 application each contain a single nucleotide error introduced by a typing mistake. The single mistake in SEQ ID NO:1 resulted in no error in the amino acid sequence of SEQ ID NO: 5. The single mistake in SEQ ID NO: 3 when translated resulted in a single amino acid error in SEQ ID NO: 7. Each of these errors is readily identifiable to one of skill in the art by cloning and sequencing the chromosomal DNA of the same microorganism used in the '667 application, which is publicly available. The identification of each of these errors is summarized in more detail below:

**(a) SEQ ID NO: 1**

By comparing the nucleotide sequence identified above as "A697-1000 Sawady [304 bp]" with the nucleotide sequence recited in the original nucleotide printout from the nucleotide sequencing machine (Exhibit 3), I confirmed that the nucleotides from positions 697-1000 in each sequence are identical. Therefore, the nucleotide at position 852 in SEQ ID NO: 1 ("G") is incorrect and should read "C." This error had no effect on the corresponding deduced amino acid sequence in SEQ ID NO:5.

**(b) SEQ ID NO: 3**

By comparing the nucleotide sequence identified above as "A"479-780 Sawady [302 bp]" with the nucleotide sequence recited in the original nucleotide printout from the nucleotide sequencing machine (Exhibit 4), I confirmed that the nucleotides from positions 479-780 in each sequence are identical. Therefore, the nucleotide at position 644 in SEQ ID NO: 3 ("A") is incorrect and should read "C."

**(c) SEQ ID NO:7**

Based on the correct nucleotide sequence for SEQ ID NO: 3, the triplet codon recited as "AAC" of nucleotide positions 643-645 in SEQ ID NO:3 should read "ACC." This should be reflected in the corresponding deduced amino acid sequence (SEQ ID NO: 7) at amino acid position 192, which was recited as "Asn" in the current

Sequence Listing. The correct codon ("ACC"), however, corresponds to the amino acid "Thr." Therefore, the amino acid at position 192 in SEQ ID NO:7 ("Asn") is incorrect and should read "Thr."

20. In sum, on the basis of the data presented herein, resequencing of the relevant parts of the chromosomal DNA of a sample of the same microorganism from which SEQ ID NOs:1 and 3 were isolated as disclosed in the '667 application (*i.e.*, *Gluconobacter oxydans* strain DSM 4025) confirms that typographical mistakes resulted in the following errors found in SEQ ID NOs: 1, 3, and 7, and that such errors would be readily identified by one skilled in this art using publicly available starting materials and routine skill. Accordingly, in my opinion, one skilled in this art would recognize, after resequencing the relevant parts of the chromosomal DNA of DSM 4025 that:

- (1) The nucleotide at position 852 of SEQ ID NO:1, which currently recites "G," should recite "C."
- (2) The nucleotide at position 644 of SEQ ID NO:3, which currently recites "A," should recite "C."
- (3) The amino acid at position 192 of SEQ ID NO:7, which currently recites "Asn," should recite "Thr."

I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Dated: September 18, 2002

Masako Shinjoh  
Masako Shinjoh

## **CURRICULUM VITAE of Masako Shinjoh**

As of August 28, 2002

**Scientist**

**Department of Applied Microbiology**

**Nippon Roche K.K.**

**Nippon Roche Research Center**

**200 Kajiwara Kamakura, Japan**

**247-8530.**

**Phone: +81-467-47-2226**

**FAX: +81-467-45-6812**

**E-mail: masako.shinjoh@roche.com**

### **Education & Research Experience:**

**1a. Scientist (April 1979 to date) at Dept. of Applied Microbiology, Nippon Roche K.K., Nippon Roche Research Center at Kamakura, Japan, which belongs to Vitamin and Fine Chemical Division in Hoffmann-La Roche**

**This work includes improvement of microorganisms producing vitamin or its precursor by conventional method and genetic engineering.**

**1b. Visiting scientist (Jan. to March 1982) at Research Institute of Molecular Biology at Nutley, NJ, USA, which belonged to Hoffmann-La Roche. Objectives: to exchange scientific information and technical transfer of genetic engineering skills.**

**2. Ph.D. (Jan. 12, 1996)**

**Ph.D. in Engineering from Department of Fermentation Technology, Osaka University, Osaka, Japan.**

**The title of the Thesis is "Metabolic engineering for 2-keto-L-gulonic acid production in**

*Gluconobacter*".

3. Master Degree (April 1977 to March 1979)

Master in Engineering from Department of Fermentation Technology, Osaka University, Osaka, Japan.

The projects involved were

"Characterization of bacteriophage of bacitracin-producing *Bacillus*"

"Application of plasmid on fermentation production: factors responsible for stabilization of hybrid plasmids carrying tryptophan operon in *E. coli*."

4. Bachelor Degree (April 1975 to March 1977)

Department of Fermentation Technology, Osaka University, Osaka, Japan.

The projects involved were

"In vitro synthesis of alpha-amylase of *Bacillus*"

5. Professional field

Microbiology

Fermentation technology

Genetic engineering

6. Memberships

a) The Society for Bioscience and Bioengineering

b) Japan Society for Bioscience, Biotechnology, and Agrochemistry

7. Personal information:

Female,

Japanese citizen,

Birthday: 20th February, 1955

## LIST OF PUBLICATIONS

### Original Papers by the Author

Shinjoh, M., Y. Setoguchi, T. Hoshino and A. Fujiwara. (1990)

L-Sorbose dissimilation in 2-keto-L-gulonic acid-producing mutant UV10 derived from *Gluconobacter melanogenus* IFO 3293. Agric. Biol. Chem. 54: 2257 - 2263.

Shinjoh, M., T. Sugisawa, S. Masuda, and T. Hoshino. (1994)

Efficient conversion of L-sorbose to 2-keto-L-gulonic acid by *Acetobacter liquefaciens* strains. J. Ferment. Bioeng. 78: 476 - 478.

Shinjoh, M., and T. Hoshino. (1995). Development of a stable shuttle vector and a conjugative transfer system for *Gluconobacter oxydans*. J. Ferment. Bioeng. 79: 95 - 99.

Shinjoh, M., N. Tomiyama, A. Asakura, and T. Hoshino. (1995) Cloning and nucleotide sequencing of membrane-bound L-sorbose dehydrogenase gene of *Acetobacter liquefaciens* IFO 12258 and its expression in *Gluconobacter oxydans*. Appl. Environ. Microbiol. 43: 1064 - 1069.

Shinjoh, M., M. Tazoe, and T. Hoshino. (2002) NADPH-dependent L-sorbose reductase is responsible for L-sorbose assimilation in *Gluconobacter suboxydans* IFO 3291. J. of Bacteriol., 84: 861 - 863.

Miyazaki, T., N. Tomiyama, M. Shinjoh, and T. Hoshino. (2002) Molecular cloning and functional expression of D-sorbitol dehydrogenase from *Gluconobacter suboxydans* IFO3255 which requires PQQ and hydrophobic protein SldB for the activity development in *E.coli*. (2001) Biosci. Biotechnol. Biochem. 66: 262-270. (the corresponding author)

Shinjoh, M., N. Tomiyama, T. Miyazaki, and T. Hoshino. (2002) Main polyol dehydrogenase of *Gluconobacter suboxydans* IFO 3255, membrane-bound D-sorbitol dehydrogenase, that needs product of upstream gene, *sldB*, for activity. Biosci. Biotechnol. Biochem. (in press)



Other Publications on the work done at Hoffmann-La Roche

Sugisawa, T., T. Hoshino, S. Masuda, S. Nomura, Y. Setoguchi, M. Tazoe, M. Shinjoh, S. Someha and A. Fujiwara. (1990) Microbial production of 2-keto-L-gulonic acid from L-sorbose and D-sorbitol by *Gluconobacter oxydans*. Agric. Biol. Chem. 54: 1201 - 1209.

Hoshino, T., T. Sugisawa, M. Tazoe, M. Shinjoh and A. Fujiwara. (1990) Metabolic pathway for 2-keto-L-gulonic acid formation in *Gluconobacter oxydans* IFO 3293. Agric. Biol. Chem. 54: 1211 - 1218.

Shinjoh, M., (1990) Biotechnology of acetic acid bacteria. Su no kagaku, Asakura shoten. Tokyo. 157 - 170. (in Japanese)

Other Publications on the work done at Osaka Univ.

Imanaka, T., K. Uchida, M. Tateishi (Shinjoh), and S. Aiba. (1979) Inducible bacteriophage of *Bacillus licheniformis* ATCC 10716. Virology 95: 249 - 250.

Tsunekawa, H., M. Tateishi (Shinjoh), T. Imanaka, S. Aiba. (1981) TnA-directed deletion of the trp operon from RSF2124-trp in *Escherichia coli*.

Patent publication: USP granted including "M. Shinjoh" as the inventor

(as of Aug. 28, 2002)

PAT. NO.	Title
1 6,407,203	Cytochrome <i>c</i> and polynucleotides encoding cytochrome <i>c</i>
2 6,146,860	Manufacture of L-ascorbic acid and D-erythorbic acid
3 6,127,156	D-sorbitol dehydrogenase gene
4 6,037,147	Cytochrome <i>c</i> and polynucleotides encoding cytochrome <i>c</i>
5 5,541,108	<i>Gluconobacter oxydans</i> strains

- 6 5,399,496 DNA shuttle vectors for *E. coli*, *Gluconobacter*, and *Acetobacter*
- 7 5,352,599 Co-enzyme-independent L-sorbose dehydrogenase of *Gluconobacter oxydans*: isolation, characterization, and cloning and autologous expression of the gene

-----END of CV-----

2<sup>nd</sup> Masako Shinojoh  
Declaration  
Exhibit - 2

## SEQUENCE LISTING

### (1) GENERAL INFORMATION

#### (i) APPLICANT

NAME: F. HOFFMANN-LA ROCHE AG  
STREET: Grezacherstrasse 124  
CITY: Basle  
COUNTRY: Switzerland  
POSTAL CODE: CH-4002  
TELEPHONE: 061 - 688 25 11  
FAX: 061 - 688 13 95  
TELEX: 962292/965542 hlr c

#### (ii) TITLE OF INVENTION:

Alcohol/Aldehyde dehydrogenase genes

#### (iii) NUMBER OF SEQUENCES: 8

#### (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: Macintosh
- (C) OPERATING SYSTEM:
- (D) SOFTWARE: MS word ver 5.1

#### (v) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION

CCAAGGCGAA	GACATGGTTT	CGAACTCGTC	GGGCCCAGTC	GTGGCAAACG	550
GCGTGATCGT	TGCCGGTTCG	ACCTGCCAAT	ACTCGCCGTT	CGGCTGCTTT	600
GTCTCGGGCC	ACGACTCGGC	CACCGGTGAA	GAGCTGTGGC	GCAACTACTT	650
CATCCCGCGC	GCTGGCGAAG	AGGGTGATGA	GACTTGGGGC	AACGATTACG	700
AAGCCCGTTG	GATGACCGGT	GCCTGGGGCC	AGATCACCTA	TGACCCCGTC	750
ACCAACCTTG	TCCACTACGG	CTCGACCGCT	GTGGGTCCGG	CGTCGGAAAC	800
CCAACGCGGC	ACCCCGGGCG	GCACGCTGTA	CGGCACGAAC	ACCCGTTTCG	850
CGGTGCGTCC	TGACACGGGC	GAGATTGTCT	GGCGTCACCA	GACCCTGCCC	900
CGCGACAAC	GGGACCAGGA	ATGCACGTTC	GAGATGATGG	TCACCAATGT	950
GGATGTCCAA	CCCTCGACCG	AGATGGAAGG	TCTGCAGTCG	ATCAACCCGA	1000
ACGCCGCAAC	TGGCGAGCGT	CGCGTGCTGA	CCGGCGTTCC	GTGCAAAACC	1050
GGCACCATGT	GGCAGTTCGA	CGCCGAAACC	GGCGAATTCC	TGTGGGCCCCG	1100
TGATACCAAC	TACCAGAACA	TGATCGAATC	CATCGACGAA	AACGGCATCG	1150
TGACCGTGAA	CGAAGATGCG	ATCCTGAAGG	AACTGGATGT	TGAATATGAC	1200
GTCTGCCCCA	CCTTCTTGGG	CGGCCGCGAC	TGGCCGTCGG	CCGCACTGAA	1250
CCCCGACAGC	GGCATCTACT	TCATCCCGCT	GAACAACGTC	TGCTATGACA	1300
TGATGGCCGT	CGATCAGGAA	TTCACCTCGA	TGGACGTCTA	TAACACCAGC	1350
AACGTGACCA	AGCTGCCGCC	CGGCAAGGAT	ATGATCGGTC	GTATTGACGC	1400
GATCGACATC	AGCACGGGTC	GTACGCTGTG	GTCGGTCGAA	CGTGCTGCGG	1450
CGAACTATTC	GCCCGTCTTG	TCGACCGGCG	GCGGCGTTCT	GTTCAACGGT	1500
GGTACGGATC	GTTACTTCCG	CGCCCTCAGC	CAAGAAACCG	GCGAGACCCT	1550
GTGGCAGACC	CGCCTTGCAA	CCGTCGCGTC	GGGCCAGGCC	ATCTCTTACG	1600
AGGTTGACGG	CATGCAATAT	GTCGCCATCG	CAGGTGGTGG	TGTCAGCTAT	1650
GGCTCGGGCC	TGAACTCGGC	ACTGGCTGGC	GAGCGAGTCG	ACTCGACCGC	1700
CATCGGTAAC	GCCGTCTACG	TCTTCGCCCT	GCCGCAATAA		1740

**INFORMATION FOR SEQ ID NO:2:**

**(i) SEQUENCE CHARACTERISTICS:**

**(A) LENGTH:** 1740 base pairs

**(B) TYPE:** nucleic acid

**(C) STRANDEDNESS:** double

**(D) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** DNA (genomic)

**(iii) ORIGINAL SOURCE:**

**ORGANISM:** *Gluconobacter oxydans*

**STRAIN:** DSM 4025

**(iv) FEATURE:**

**FEATURED KEY:** CDS

**POSITION:** 1..1737

**SEQUENCING METHOD:** E

ATGAAGACGT	CGTCTTTGCT	GGTTGCGAGC	GTTGCCGCGC	TTGCAAGCTA	50
TAGCTCCTTT	GCGCTTGCTC	AAGTGACCCC	CGTCACCGAT	GAATTGCTGG	100
CGAACCCGCC	CGCTGGTGAA	TGGATCAGCT	ACGGTCAGAA	CCAAGAAAAC	150
TACCGTCACT	CGCCCCTGAC	GCAGATCACG	ACTGAGAACG	TCGGCCAACT	200
GCAACTGGTC	TGGGCGCGCG	GCATGCAGCC	GGGCAAAGTC	CAAGTCACGC	250
CCCTGATCCA	TGACGGCGTC	ATGTATCTGG	CAAACCCGGG	CGACGTGATC	300

CAGGCCATCG ACGCCAAAAC TGGCGATCTG ATCTGGGAAC ACCGCCGCCA 350  
ACTGCCGAAC ATCGCCACGC TGAACAGCTT TGGCGAGCCG ACCCGCGGCA 400  
TGGCGCTGTA CGGCACCAAC GTTTACTTTG TTTCGTGGGA CAACCACCTG 450  
GTCGCCCTCG ACACCGCAAC TGGCCAAGTG ACGTTCGACG TCGACCGCGG 500  
CCAAGGCGAA GACATGGTTT CGAACTCGTC GGGCCCGATC GTGGCAAACG 550  
GCGTGATCGT TGCCGGTTCG ACCTGCCAAT ACTCGCCGTT CGGCTGCTTT 600  
GTCTCGGGCC ACGACTCGGC CACCGGTGAA GAGCTGTGGC GCAACTACTT 650  
CATCCCGCGC GCTGGCGAAG AGGGTGATGA GACTTGGGGC AACGATTACG 700  
AAGCCCGTTG GATGACCGGC GTCTGGGGTC AGATCACCTA TGACCCCGTT 750  
GGCGGCCTTG TCCACTACGG CTCGTCGGCT GTTGGCCCGG CTTCGGAAAC 800  
CCAGCGCGGC ACCACCGGCG GCACCATGTA CGGCACCAAC ACCCGTTTCG 850  
CTGTCCGTCC CGAGACTGGC GAGATCGTCT GGCGTCACCA AACTCTGCCC 900  
CGCGACAAC TGGACCAAGA GTGCACCTTC GAGATGATGG TTGCCAACGT 950  
TGACGTGCAG CCCGCAGCTG ACATGGACGG CGTCCGCTCG ATCAACCCGA 1000  
ACGCCGCCAC CGGCGAGCGT CGCGTTCTGA CCGGCGTTCC GTGCAAAACC 1050  
GGCACCATGT GGCAGTTCGA CGCCGAAACC GGCGAATTCC TGTGGGCCCCG 1100  
TGACACCAGC TACGAGAACA TCATCGAATC GATCGACGAA AACGGCATCG 1150  
TGACCGTCGA CGAGTCGAAA GTTCTGACCG AGCTGGACAC CCCCTATGAC 1200  
GTCTGCCCCG TGCTGCTGGG TGGCCGTGAC TGGCCGTCGG CTGCGCTGAA 1250  
CCCCGATACC GGCATCTACT TTATCCCGCT GAACAACACC TGCATGGATA 1300  
TCGAAGCTGT CGACCAGGAA TTCAGCTCGC TGGACGTGTA CAACCAAAGC 1350  
CTGACCGCCA AAATGGCACC GGGTAAAGAG CTGGTTGGCC GTATCGACGC 1400  
CATCGACATC AGCACAGGCC GCACCCTGTG GACCGCTGAG CGCGAAGCCT 1450  
CGAACTACGC GCCTGTCCTG TCGACCGCTG GCGGCGTTCT GTTCAACGGC 1500  
GGCACCACC GTTACTTCCG CGCTCTCAGC CAAGAGACCG GCGAGACCCT 1550

GTGGCAGACC CGTCTGGCGA CTGTCGCTTC GGGCCAAGCT GTCTCGTACG 1600  
 AGATCGACGG CGTCCAATAC ATCGCCATCG GCGGCGGCGG CACGACCTAT 1650  
 GGTTCGTTCC ACAACCGTCC CCTGGCCGAG CCGGTCGACT CGACCGCGAT 1700  
 CGGTAATGCG ATGTACGTCT TCGCGCTGCC CCAGCAATAA 1740

**INFORMATION FOR SEQ ID NO:3:**

**(i) SEQUENCE CHARACTERISTICS:**

**(A) LENGTH:** 1737 base pairs

**(B) TYPE:** nucleic acid

**(C) STRANDEDNESS:** double

**(D) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** DNA (genomic)

**(iii) ORIGINAL SOURCE:**

**ORGANISM:** *Gluconobacter oxydans*

**STRAIN:** DSM 4025

**(iv) FEATURE:**

**FEATURE KEY:** CDS

**POSITION:** 1..1734

**SEQUENCING METHOD:** E

ATGAAACTGA CGACCCTGCT GCAAAGCAGC GCCGCCCTGC TTGTGCTTGG 50  
 CACCATTTCCC GCCCTTGCCC AAACCGCCAT CACCGATGAA ATGCTGGCGA 100

ACCCGCCCCG	TGGTGAATGG	ATCAACTACG	GTCAGAACCA	AGAGAACTAC	150
CGCCACTCGC	CCCTGACGCA	GATTACCGCA	GACAACGTCG	GCCAACTGCA	200
ACTGGTCTGG	GCGCGCGGTA	TGGAAGCGGG	CAAGATCCAA	GTGACCCCGC	250
TTGTCCATGA	CGGCGTCATG	TATCTGGCAA	ACCCCGGTGA	CGTGATCCAG	300
GCCATCGACG	CCGCGACCGG	CGATCTGATC	TGGGAACACC	GCCGCCAACT	350
GCCGAACATC	GCCACGCTGA	ACAGCTTTGG	TGAGCCGACC	CGCGGCATGG	400
CCCTCTATGG	CACCAACGTC	TATTTTCGTCT	CGTGGGACAA	CCACTTGGTC	450
GCGCTGGACA	CCTCGACCGG	CCAAGTCGTA	TTCGACGTCG	ATCGCGGTCA	500
AGGCACGGAT	ATGGTCTCGA	ACTCGTCCGG	CCCGATTGTC	GCCAATGGCG	550
TCATCGTTGC	GGGCTCGACC	TGTCAGTATT	CGCCGTTTCGG	CTGTTTCGTT	600
TCGGGCCACG	ACTCGGCCAC	CGGTGAAGAG	CTGTGGCGCA	ACAACTTTAT	650
CCCGCGCGCC	GGCGAAGAGG	GTGATGAGAC	CTGGGGCAAT	GATTACGAGG	700
CCCGCTGGAT	GACCGGCGTT	TGGGGCCAGA	TCACCTATGA	CCCGTTGGC	750
GGCCTTGTCC	ACTACGGCAC	CTCAGCAGTT	GGCCCTGCGG	CCGAGATTCA	800
GCGCGGCACC	GTTGGCGGCT	CGATGTATGG	CACCAACACC	CGCTTTGCTG	850
TCCGCCCCGA	GACCGGCGAG	ATCGTCTGGC	GTCACCAAAC	TCTGCCCCGC	900
GACAACTGGG	ACCAAGAGTG	TACGTTTCGAG	ATGATGGTCG	TCAACGTCGA	950
CGTCCAGCCC	TCGGCTGAGA	TGGAAGGCCT	GCACGCCATC	AACCCCGATG	1000
CGCCACGGG	CGAGCGTCGC	GTTGTGACCG	GCGTTCCGTG	CAAGAACGGC	1050
ACCATGTGGC	AGTTCGACGC	CGAAACCGGC	GAATTCCTGT	GGGCGCGCGA	1100
CACCAGCTAT	CAGAACCTGA	TCGAAAGCGT	CGATCCCGAT	GGTCTGGTGC	1150
ATGTGAACGA	AGATCTGGTC	GTGACCGAGC	TGGAAGTGGC	CTATGAAATC	1200
TGCCCCGACCT	TCCTGGGTGG	CCGCGACTGG	CCGTCGGCTG	CGCTGAACCC	1250
CGATACTGGC	ATCTATTTCA	TCCCGCTGAA	CAACGCCTGT	AGCGGTATGA	1300
CGGCTGTCTGA	CCAAGAGTTC	AGCTCGCTCG	ATGTGTATAA	CGTCAGCCTC	1350



GACTATAAAC TGTCGCCCCG TTCGGAAAAC ATGGGCCGTA TCGACGCCAT 1400  
 CGACATCAGC ACCGGCCGCA CGCTGTGGTC GGCTGAACGC TACGCCTCGA 1450  
 ACTACGCGCC TGTCCTGTCC ACCGGCGGCG GCGTGCTGTT CAACGGCGGC 1500  
 ACCGACCGTT ACTTCCGCGC CCTCAGCCAA GAGACCGGCG AGACGCTGTG 1550  
 GCAGACCCGT CTGGCGACTG TCGCCTCGGG TCAAGCGATT TCCTATGAGA 1600  
 TCGACGGCGT GCAATATGTC GCCATCGGGC GCGGCGGCAC CAGCTATGGC 1650  
 AGCAACCACA ACCGCGCCCT GACCGAGCGG ATCGACTCGA CCGCCATCGG 1700  
 CAGCGCGATC TATGTCTTTG CTCTGCCGCA GCAGTAA 1737

#### INFORMATION FOR SEQ ID NO:4:

(i) **SEQUENCE CHARACTERISTICS:**

(A) **LENGTH:** 1740 base pairs

(B) **TYPE:** nucleic acid

(C) **STRANDEDNESS:** double

(D) **TOPOLOGY:** linear

(ii) **MOLECULE TYPE:** DNA (genomic)

(iii) **ORIGINAL SOURCE:**

**ORGANISM:** *Gluconobacter oxydans*

**STRAIN:** DSM 4025

(iv) **FEATURE:**

**FEATURE KEY:** CDS

**POSITION:** 1..1737

**SEQUENCING METHOD:** E

ATGAACCCCA	CAACGCTGCT	TCGCACCAGC	GCGGCCGTGC	TATTGCTTAC	50
CGCGCCCGCC	GCATTCGCGC	AGGTAACCCC	GATTACCGAT	GAACTGCTGG	100
CGAACCCGCC	CGCTGGTGAA	TGGATTAACT	ACGGCCGCAA	CCAAGAAAAC	150
TATCGCCACT	CGCCCCTGAC	CCAGATCACT	GCCGACAACG	TTGGTCAGTT	200
GCAACTGGTC	TGGGCCCGCG	GGATGGAGGC	GGGGGCCGTA	CAGGTCACGC	250
CGATGATCCA	TGATGGCGTG	ATGTATCTGG	CAAACCCCGG	TGATGTGATC	300
CAGGCGCTGG	ATGCGCAAAC	AGGCGATCTG	ATCTGGGAAC	ACCGCCGCCA	350
ACTGCCCGCC	GTCGCCACGC	TAAACGCCCA	AGGCGACCGC	AAGCGCGGCG	400
TCGCCCTTTA	CGGCACGAGC	CTCTATTTCA	GCTCATGGGA	CAACCATCTG	450
ATCGCGCTGG	ATATGGAGAC	GGGCCAGGTC	GTATTTCGATG	TCGAACGTGG	500
ATCGGGCGAA	GACGGCTTGA	CCAGTAACAC	CACGGGGCCG	ATTGTCGCCA	550
ATGGCGTCAT	CGTCGCGGGT	TCCACCTGCC	AATATTCGCC	CTATGGATGC	600
TTTATCTCGG	GGCACGATTC	CGCGACGGGT	GAGGAGCTGT	GGCGCAACCA	650
CTTTATCCCG	CAGCCGGGCG	AAGAGGGTGA	CGAGACTTGG	GGCAATGATT	700
TCGAGGCGCG	CTGGATGACC	GGCGTCTGGG	GTCAGATCAC	CTATGATCCC	750
GTGACGAACC	TTGTGTTCTA	TGGCTCGACC	GGCGTGGGCC	CAGCGTCCGA	800
AACCCAGCGC	GGCACGCCGG	GCGGCACGCT	GTATGGCACC	AACACCCGCT	850
TTGCGGTGCG	TCCCGACACG	GGCGAGATTG	TCTGGCGTCA	CCAGACCCTG	900
CCGCGCGACA	ACTGGGACCA	AGAATGCACG	TTGAGATGA	TGGTCGCCAA	950
CGTCGATGTG	CAACCCTCGG	CCGAGATGGA	GGGTCTGCGC	GCCATCAACC	1000
CCAATGCGGC	GACGGGCGAG	CGCCGTGTGC	TGACGGGTGC	GCCTTGCAAG	1050
ACCGGCACGA	TGTGGTCGTT	TGATGCGGCC	TCGGGCGAAT	TCCTGTGGGC	1100
GCGTGATACC	AACTACACCA	ATATGATCGC	CTCGATCGAC	GAGACCGGCC	1150
TTGTGACGGT	GAACGAGGAT	GCGGTGCTGA	AAGAGCTGGA	CGTTGAATAT	1200

GACGTCTGCC CGACCTTCCT GGGTGGGCGC GACTGGTCGT CAGCCGCACT 1250  
 GAACCCGGAC ACCGGCATT TTT ACTTCTTGCC GCTGAACAAT GCCTGCTACG 1300  
 ATATTATGGC CGTTGATCAA GAGTTTAGCG CGCTCGACGT CTATAACACC 1350  
 AGCGCGACCG CAAAACTCGC GCCGGGCTTT GAAAATATGG GCCGCATCGA 1400  
 CGCGATTGAT ATCAGCACCG GGCGCACCTT GTGGTCGGCG GAGCGCCCTG 1450  
 CGGCGAACTA CTCGCCCCGTT TTGTCGACGG CAGGCGGTGT GGTGTTCAAC 1500  
 GGCGGGACCG ACCGCTATTT CCGTGCCCTC AGCCAGGAAA CCGGCGAGAC 1550  
 TTTGTGGCAG GCCCGTCTTG CGACGGTCGC GACGGGGCAG GCGATCAGCT 1600  
 ACGAGTTGGA CGGCGTGCAA TATATCGCCA TCGGTGCGGG CCGTCTGACC 1650  
 TATGGCACGC AATTGAACGC GCCGCTGGCC GAGGCAATCG ATTCGACCTC 1700  
 GGTCGGTAAT GCGATCTATG TCTTTGCACT GCCGCAGTAA 1740

# INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 579 residues

(B) TYPE: amino acid

(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) ORIGINAL SOURCE:

ORGANISM: *Gluconobacter oxydans*

STRAIN: DSM 4025

(iv) FEATURE:

FEATURE KEY: sig peptide

POSITION: -23..-1

SEQUENCING METHOD: E

FEATURE KEY: mat peptide

POSITION: 1..556

SEQUENCING METHOD: E

Met Lys Pro Thr Ser Leu Leu Trp Ala Ser Ala Gly Ala Leu Ala  
                  -20                  -15                  -10  
Leu Leu Ala Ala Pro Ala Phe Ala Gln Val Thr Pro Val Thr Asp  
                  -5                  1                  5  
Glu Leu Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Ser Tyr Gly  
          10                  15                  20  
Gln Asn Gln Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr  
          25                  30                  35  
Thr Glu Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met  
          40                  45                  50  
Gln Pro Gly Lys Val Gln Val Thr Pro Leu Ile His Asp Gly Val  
          55                  60                  65  
Met Tyr Leu Ala Asn Pro Gly Asp Val Ile Gln Ala Ile Asp Ala  
          70                  75                  80  
Lys Thr Gly Asp Leu Ile Trp Glu His Arg Arg Gln Leu Pro Asn  
          85                  90                  95  
Ile Ala Thr Leu Asn Ser Phe Gly Glu Pro Thr Arg Gly Met Ala  
          100                  105                  110  
Leu Tyr Gly Thr Asn Val Tyr Phe Val Ser Trp Asp Asn His Leu  
          115                  120                  125  
Val Ala Leu Asp Thr Ala Thr Gly Gln Val Thr Phe Asp Val Asp  
          130                  135                  140

Arg Gly Gln Gly Glu Asp	Met Val Ser Asn Ser Ser Gly Pro Ile
145	150 155
Val Ala Asn Gly Val Ile Val	Ala Gly Ser Thr Cys Gln Tyr Ser
160	165 170
Pro Phe Gly Cys Phe Val	Ser Gly His Asp Ser Ala Thr Gly Glu
175	180 185
Glu Leu Trp Arg Asn Tyr	Phe Ile Pro Arg Ala Gly Glu Glu Gly
190	195 200
Asp Glu Thr Trp Gly Asn	Asp Tyr Glu Ala Arg Trp Met Thr Gly
205	210 215
Ala Trp Gly Gln Ile Thr	Tyr Asp Pro Val Thr Asn Leu Val His
220	225 230
Tyr Gly Ser Thr Ala Val	Gly Pro Ala Ser Glu Thr Gln Arg Gly
235	240 245
Thr Pro Gly Gly Thr Leu	Tyr Gly Thr Asn Thr Arg Phe Ala Val
250	255 260
Arg Pro Asp Thr Gly Glu	Ile Val Trp Arg His Gln Thr Leu Pro
265	270 275
Arg Asp Asn Trp Asp Gln	Glu Cys Thr Phe Glu Met Met Val Thr
280	285 290
Asn Val Asp Val Gln Pro	Ser Thr Glu Met Glu Gly Leu Gln Ser
295	300 305
Ile Asn Pro Asn Ala Ala	Thr Gly Glu Arg Arg Val Leu Thr Gly
310	315 320
Val Pro Cys Lys Thr Gly	Thr Met Trp Gln Phe Asp Ala Glu Thr
325	330 335
Gly Glu Phe Leu Trp Ala	Arg Asp Thr Asn Tyr Gln Asn Met Ile
340	345 350
Glu Ser Ile Asp Glu Asn	Gly Ile Val Thr Val Asn Glu Asp Ala
355	360 365
Ile Leu Lys Glu Leu Asp	Val Glu Tyr Asp Val Cys Pro Thr Phe
370	375 380
Leu Gly Gly Arg Asp Trp	Pro Ser Ala Ala Leu Asn Pro Asp Ser

385	390	395
Gly Ile Tyr Phe Ile Pro	Leu Asn Asn Val	Cys Tyr Asp Met Met
400	405	410
Ala Val Asp Gln Glu Phe Thr	Ser Met Asp	Val Tyr Asn Thr Ser
415	420	425
Asn Val Thr Lys Leu Pro	Pro Gly Lys Asp	Met Ile Gly Arg Ile
430	435	440
Asp Ala Ile Asp Ile Ser Thr	Gly Arg Thr	Leu Trp Ser Val Glu
445	450	455
Arg Ala Ala Ala Asn Tyr Ser	Pro Val Leu	Ser Thr Gly Gly Gly
460	465	470
Val Leu Phe Asn Gly Gly Thr	Asp Arg Tyr	Phe Arg Ala Leu Ser
475	480	485
Gln Glu Thr Gly Glu Thr Leu	Trp Gln Thr	Arg Leu Ala Thr Val
490	495	500
Ala Ser Gly Gln Ala Ile Ser	Tyr Glu Val	Asp Gly Met Gln Tyr
505	510	515
Val Ala Ile Ala Gly Gly Gly	Val Ser Tyr	Gly Ser Gly Leu Asn
520	525	530
Ser Ala Leu Ala Gly Glu Arg	Val Asp Ser	Thr Ala Ile Gly Asn
535	540	545
Ala Val Tyr Val Phe Ala Leu	Pro Gln	
550	555	

INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 579 residues

(B) TYPE: amino acid  
 (C) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: protein  
 (iii) ORIGINAL SOURCE:  
     ORGANISM: *Gluconobacter oxydans*  
     STRAIN: DSM 4025  
 (iv) FEATURE:

FEATURE KEY: sig peptide

POSITION: -23..-1

SEQUENCING METHOD: S

FEATURE KEY: mat peptide

POSITION: 1..556

SEQUENCING METHOD: S

Met Lys Thr Ser Ser Leu Leu Val Ala Ser Val Ala Ala Leu Ala  
           -20                    -15                    -10  
 Ser Tyr Ser Ser Phe Ala Leu Ala Gln Val Thr Pro Val Thr Asp  
           -5                            1                    5  
 Glu Leu Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Ser Tyr Gly  
           10                            15                    20  
 Gln Asn Gln Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr  
           25                            30                    35  
 Thr Glu Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met  
           40                            45                    50  
 Gln Pro Gly Lys Val Gln Val Thr Pro Leu Ile His Asp Gly Val  
           55                            60                    65

Met	Tyr	Leu	Ala	Asn	Pro	Gly	Asp	Val	Ile	Gln	Ala	Ile	Asp	Ala	70	75	80
Lys	Thr	Gly	Asp	Leu	Ile	Trp	Glu	His	Arg	Arg	Gln	Leu	Pro	Asn	85	90	95
Ile	Ala	Thr	Leu	Asn	Ser	Phe	Gly	Glu	Pro	Thr	Arg	Gly	Met	Ala	100	105	110
Leu	Tyr	Gly	Thr	Asn	Val	Tyr	Phe	Val	Ser	Trp	Asp	Asn	His	Leu	115	120	125
Val	Ala	Leu	Asp	Thr	Ala	Thr	Gly	Gln	Val	Thr	Phe	Asp	Val	Asp	130	135	140
Arg	Gly	Gln	Gly	Glu	Asp	Met	Val	Ser	Asn	Ser	Ser	Gly	Pro	Ile	145	150	155
Val	Ala	Asn	Gly	Val	Ile	Val	Ala	Gly	Ser	Thr	Cys	Gln	Tyr	Ser	160	165	170
Pro	Phe	Gly	Cys	Phe	Val	Ser	Gly	His	Asp	Ser	Ala	Thr	Gly	Glu	175	180	185
Glu	Leu	Trp	Arg	Asn	Tyr	Phe	Ile	Pro	Arg	Ala	Gly	Glu	Glu	Gly	190	195	200
Asp	Glu	Thr	Trp	Gly	Asn	Asp	Tyr	Glu	Ala	Arg	Trp	Met	Thr	Gly	205	210	215
Val	Trp	Gly	Gln	Ile	Thr	Tyr	Asp	Pro	Val	Gly	Gly	Leu	Val	His	220	225	230
Tyr	Gly	Ser	Ser	Ala	Val	Gly	Pro	Ala	Ser	Glu	Thr	Gln	Arg	Gly	235	240	245
Thr	Thr	Gly	Gly	Thr	Met	Tyr	Gly	Thr	Asn	Thr	Arg	Phe	Ala	Val	250	255	260
Arg	Pro	Glu	Thr	Gly	Glu	Ile	Val	Trp	Arg	His	Gln	Thr	Leu	Pro	265	270	275
Arg	Asp	Asn	Trp	Asp	Gln	Glu	Cys	Thr	Phe	Glu	Met	Met	Val	Ala	280	285	290
Asn	Val	Asp	Val	Gln	Pro	Ala	Ala	Asp	Met	Asp	Gly	Val	Arg	Ser	295	300	305
Ile	Asn	Pro	Asn	Ala	Ala	Thr	Gly	Glu	Arg	Arg	Val	Leu	Thr	Gly			



310	315	320
Val Pro Cys Lys Thr Gly Thr Met Trp Gln Phe Asp Ala Glu Thr		
325	330	335
Gly Glu Phe Leu Trp Ala Arg Asp Thr Ser Tyr Glu Asn Ile Ile		
340	345	350
Glu Ser Ile Asp Glu Asn Gly Ile Val Thr Val Asp Glu Ser Lys		
355	360	365
Val Leu Thr Glu Leu Asp Thr Pro Tyr Asp Val Cys Pro Leu Leu		
370	375	380
Leu Gly Gly Arg Asp Trp Pro Ser Ala Ala Leu Asn Pro Asp Thr		
385	390	395
Gly Ile Tyr Phe Ile Pro Leu Asn Asn Thr Cys Met Asp Ile Glu		
400	405	410
Ala Val Asp Gln Glu Phe Ser Ser Leu Asp Val Tyr Asn Gln Ser		
415	420	425
Leu Thr Ala Lys Met Ala Pro Gly Lys Glu Leu Val Gly Arg Ile		
430	435	440
Asp Ala Ile Asp Ile Ser Thr Gly Arg Thr Leu Trp Thr Ala Glu		
445	450	455
Arg Glu Ala Ser Asn Tyr Ala Pro Val Leu Ser Thr Ala Gly Gly		
460	465	470
Val Leu Phe Asn Gly Gly Thr Asp Arg Tyr Phe Arg Ala Leu Ser		
475	480	485
Gln Glu Thr Gly Glu Thr Leu Trp Gln Thr Arg Leu Ala Thr Val		
490	495	500
Ala Ser Gly Gln Ala Val Ser Tyr Glu Ile Asp Gly Val Gln Tyr		
505	510	515
Ile Ala Ile Gly Gly Gly Gly Thr Thr Tyr Gly Ser Phe His Asn		
520	525	530
Arg Pro Leu Ala Glu Pro Val Asp Ser Thr Ala Ile Gly Asn Ala		
535	540	545
Met Tyr Val Phe Ala Leu Pro Gln Gln		
550	555	

**INFORMATION FOR SEQ ID NO:7:**

**(i) SEQUENCE CHARACTERISTICS:**

**(A) LENGTH:** 578 residues

**(B) TYPE:** amino acid

**(C) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** protein

**(iii) ORIGINAL SOURCE:**

**ORGANISM:** *Gluconobacter oxydans*

**STRAIN:** DSM 4025

**(iv) FEATURE:**

**FEATURE KEY:** sig peptide

**POSITION:** -23...-1

**SEQUENCING METHOD:** S

**FEATURE KEY:** mat peptide

**POSITION:** 1..555

**SEQUENCING METHOD:** S

Met Lys Leu Thr Thr Leu Leu Gln Ser Ser Ala Ala Leu Leu Val  
-20 -15 -10

Leu Gly Thr Ile Pro Ala Leu Ala Gln Thr Ala Ile Thr Asp Glu

-5	1	5	
Met Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Asn Tyr Gly Gln	10	15	20
Asn Gln Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr Ala	25	30	35
Asp Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met Glu	40	45	50
Ala Gly Lys Ile Gln Val Thr Pro Leu Val His Asp Gly Val Met	55	60	65
Tyr Leu Ala Asn Pro Gly Asp Val Ile Gln Ala Ile Asp Ala Ala	70	75	80
Thr Gly Asp Leu Ile Trp Glu His Arg Arg Gln Leu Pro Asn Ile	85	90	95
Ala Thr Leu Asn Ser Phe Gly Glu Pro Thr Arg Gly Met Ala Leu	100	105	110
Tyr Gly Thr Asn Val Tyr Phe Val Ser Trp Asp Asn His Leu Val	115	120	125
Ala Leu Asp Thr Ser Thr Gly Gln Val Val Phe Asp Val Asp Arg	130	135	140
Gly Gln Gly Thr Asp Met Val Ser Asn Ser Ser Gly Pro Ile Val	145	150	155
Ala Asn Gly Val Ile Val Ala Gly Ser Thr Cys Gln Tyr Ser Pro	160	165	170
Phe Gly Cys Phe Val Ser Gly His Asp Ser Ala Thr Gly Glu Glu	175	180	185
Leu Trp Arg Asn Asn Phe Ile Pro Arg Ala Gly Glu Glu Gly Asp	190	195	200
Glu Thr Trp Gly Asn Asp Tyr Glu Ala Arg Trp Met Thr Gly Val	205	210	215
Trp Gly Gln Ile Thr Tyr Asp Pro Val Gly Gly Leu Val His Tyr	220	225	230
Gly Thr Ser Ala Val Gly Pro Ala Ala Glu Ile Gln Arg Gly Thr	235	240	245

Val Gly Gly Ser Met Tyr Gly Thr Asn Thr Arg Phe Ala Val Arg  
 250 255 260  
 Pro Glu Thr Gly Glu Ile Val Trp Arg His Gln Thr Leu Pro Arg  
 265 270 275  
 Asp Asn Trp Asp Gln Glu Cys Thr Phe Glu Met Met Val Val Asn  
 280 285 290  
 Val Asp Val Gln Pro Ser Ala Glu Met Glu Gly Leu His Ala Ile  
 295 300 305  
 Asn Pro Asp Ala Ala Thr Gly Glu Arg Arg Val Val Thr Gly Val  
 310 315 320  
 Pro Cys Lys Asn Gly Thr Met Trp Gln Phe Asp Ala Glu Thr Gly  
 325 330 335  
 Glu Phe Leu Trp Ala Arg Asp Thr Ser Tyr Gln Asn Leu Ile Glu  
 340 345 350  
 Ser Val Asp Pro Asp Gly Leu Val His Val Asn Glu Asp Leu Val  
 355 360 365  
 Val Thr Glu Leu Glu Val Ala Tyr Glu Ile Cys Pro Thr Phe Leu  
 370 375 380  
 Gly Gly Arg Asp Trp Pro Ser Ala Ala Leu Asn Pro Asp Thr Gly  
 385 390 395  
 Ile Tyr Phe Ile Pro Leu Asn Asn Ala Cys Ser Gly Met Thr Ala  
 400 405 410  
 Val Asp Gln Glu Phe Ser Ser Leu Asp Val Tyr Asn Val Ser Leu  
 415 420 425  
 Asp Tyr Lys Leu Ser Pro Gly Ser Glu Asn Met Gly Arg Ile Asp  
 430 435 440  
 Ala Ile Asp Ile Ser Thr Gly Arg Thr Leu Trp Ser Ala Glu Arg  
 445 450 455  
 Tyr Ala Ser Asn Tyr Ala Pro Val Leu Ser Thr Gly Gly Gly Val  
 460 465 470  
 Leu Phe Asn Gly Gly Thr Asp Arg Tyr Phe Arg Ala Leu Ser Gln  
 475 480 485  
 Glu Thr Gly Glu Thr Leu Trp Gln Thr Arg Leu Ala Thr Val Ala  
 490 495 500

Ser Gly Gln Ala Ile Ser Tyr Glu Ile Asp Gly Val Gln Tyr Val  
 505 510 515  
 Ala Ile Gly Arg Gly Gly Thr Ser Tyr Gly Ser Asn His Asn Arg  
 520 525 530  
 Ala Leu Thr Glu Arg Ile Asp Ser Thr Ala Ile Gly Ser Ala Ile  
 535 540 545  
 Tyr Val Phe Ala Leu Pro Gln Gln  
 550 555

**INFORMATION FOR SEQ ID NO:8:**

**(i) SEQUENCE CHARACTERISTICS:**

**(A) LENGTH:** 579 residues

**(B) TYPE:** amino acid

**(C) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** protein

**(iii) ORIGINAL SOURCE:**

**ORGANISM:** *Gluconobacter oxydans*

**STRAIN:** DSM 4025

**(iv) FEATURE:**

**FEATURE KEY:** sig peptide

**POSITION:** -23..-1

**SEQUENCING METHOD:** E

FEATURE KEY: mat peptide

POSITION: 1..556

SEQUENCING METHOD: E

Met Asn Pro Thr Thr Leu Leu Arg Thr Ser Ala Ala Val Leu Leu  
-20 -15 -10  
Leu Thr Ala Pro Ala Ala Phe Ala Gln Val Thr Pro Ile Thr Asp  
-5 1 5  
Glu Leu Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Asn Tyr Gly  
10 15 20  
Arg Asn Gln Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr  
25 30 35  
Ala Asp Asn Val Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met  
40 45 50  
Glu Ala Gly Ala Val Gln Val Thr Pro Met Ile His Asp Gly Val  
55 60 65  
Met Tyr Leu Ala Asn Pro Gly Asp Val Ile Gln Ala Leu Asp Ala  
70 75 80  
Gln Thr Gly Asp Leu Ile Trp Glu His Arg Arg Gln Leu Pro Ala  
85 90 95  
Val Ala Thr Leu Asn Ala Gln Gly Asp Arg Lys Arg Gly Val Ala  
100 105 110  
Leu Tyr Gly Thr Ser Leu Tyr Phe Ser Ser Trp Asp Asn His Leu  
115 120 125  
Ile Ala Leu Asp Met Glu Thr Gly Gln Val Val Phe Asp Val Glu  
130 135 140  
Arg Gly Ser Gly Glu Asp Gly Leu Thr Ser Asn Thr Thr Gly Pro  
145 150 155  
Ile Val Ala Asn Gly Val Ile Val Ala Gly Ser Thr Cys Gln Tyr  
160 165 170  
Ser Pro Tyr Gly Cys Phe Ile Ser Gly His Asp Ser Ala Thr Gly

175	180	185
Glu Glu Leu Trp Arg Asn His 190	Phe Ile Pro Gln Pro Gly Glu Glu 195	
Gly Asp Glu Thr Trp Gly Asn 205	Asp Phe Glu Ala Arg Trp Met Thr 210	
Gly Val Trp Gly Gln Ile Thr 220	Tyr Asp Pro Val Thr Asn Leu Val 225	
Phe Tyr Gly Ser Thr Gly Val 235	Gly Pro Ala Ser Glu Thr Gln Arg 240	
Gly Thr Pro Gly Gly Thr Leu 250	Tyr Gly Thr Asn Thr Arg Phe Ala 255	
Val Arg Pro Asp Thr Gly Glu 265	Ile Val Trp Arg His Gln Thr Leu 270	
Pro Arg Asp Asn Trp Asp Gln 280	Glu Cys Thr Phe Glu Met Met Val 285	
Ala Asn Val Asp Val Gln Pro 295	Ser Ala Glu Met Glu Gly Leu Arg 300	
Ala Ile Asn Pro Asn Ala Ala 310	Thr Gly Glu Arg Arg Val Leu Thr 315	
Gly Ala Pro Cys Lys Thr Gly 325	Thr Met Trp Ser Phe Asp Ala Ala 330	
Ser Gly Glu Phe Leu Trp Ala 340	Arg Asp Thr Asn Tyr Thr Asn Met 345	
Ile Ala Ser Ile Asp Glu Thr 355	Gly Leu Val Thr Val Asn Glu Asp 360	
Ala Val Leu Lys Glu Leu Asp 370	Val Glu Tyr Asp Val Cys Pro Thr 375	
Phe Leu Gly Gly Arg Asp Trp 385	Ser Ser Ala Ala Leu Asn Pro Asp 390	
Thr Gly Ile Tyr Phe Leu Pro 400	Leu Asn Asn Ala Cys Tyr Asp Ile 405	
Met Ala Val Asp Gln Glu Phe 415	Ser Ala Leu Asp Val Tyr Asn Thr 420	

Ser	Ala	Thr	Ala	Lys	Leu	Ala	Pro	Gly	Phe	Glu	Asn	Met	Gly	Arg
430						435				440				
Ile	Asp	Ala	Ile	Asp	Ile	Ser	Thr	Gly	Arg	Thr	Leu	Trp	Ser	Ala
445						450				455				
Glu	Arg	Pro	Ala	Ala	Asn	Tyr	Ser	Pro	Val	Leu	Ser	Thr	Ala	Gly
460						465				470				
Gly	Val	Val	Phe	Asn	Gly	Gly	Thr	Asp	Arg	Tyr	Phe	Arg	Ala	Leu
475						480				485				
Ser	Gln	Glu	Thr	Gly	Glu	Thr	Leu	Trp	Gln	Ala	Arg	Leu	Ala	Thr
490						495				500				
Val	Ala	Thr	Gly	Gln	Ala	Ile	Ser	Tyr	Glu	Leu	Asp	Gly	Val	Gln
505						510				515				
Tyr	Ile	Ala	Ile	Gly	Ala	Gly	Gly	Leu	Thr	Tyr	Gly	Thr	Gln	Leu
520						525				530				
Asn	Ala	Pro	Leu	Ala	Glu	Ala	Ile	Asp	Ser	Thr	Ser	Val	Gly	Asn
535						540				545				
Ala	Ile	Tyr	Val	Phe	Ala	Leu	Pro	Gln						
550						555								



**INFORMATION FOR SEQ ID NO:9:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH:** 82 bases
- (B) TYPE:** nucleotide
- (C) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** DNA

**(iii) ORIGINAL SOURCE:** synthetic oligonucleotide

CATGAAAATA AAAACAGGTG CACGCATCCT CGCATTATCC GCATTAACGA 50  
CGATGATGTT TTCCGCCTCG GCTCTCGCCC AG 82

**INFORMATION FOR SEQ ID NO:10:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH:** 83 bases
- (B) TYPE:** nucleotide
- (C) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** DNA

**(iii) ORIGINAL SOURCE:** synthetic oligonucleotide

GTTACCTGGG CGAGAGCCGA GCGGAAAAC ATCATCGTCG TTAATGCGGA 50  
TAATGCGAGG ATGCGTGCAC CTGTTTTTAT TTT 83

**INFORMATION FOR SEQ ID NO:11:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH:** 27 residues
- (B) TYPE:** amino acid
- (C) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** peptide

**(iii) ORIGINAL SOURCE:** *E. coli*

**(iv) FEATURE:**

**FEATURE KEY:** sig peptide

**POSITION:** 1..26

**FEATURE METHOD:** S

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu  
1 5 10 15  
Thr Thr Met Met Phe Ser Ala Ser Ala Leu Ala Gln  
20 25 27

**INFORMATION FOR SEQ ID NO:12:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH:** 27 bases
- (B) TYPE:** nucleotide
- (C) TOPOLOGY:** linear

**(ii) MOLECULE TYPE:** DNA

**(iii) ORIGINAL SOURCE:** synthetic oligonucleotide

GTTAGCGCGG TGGATCCCCA TTGGAGG

27

\*\*\* GENIAS - Nucleotide Usage Table \*\*\*  
 DNA Sequence Name (1): pEnza (1-1740)  
 Comment :

T	319	( 18.3%)
A	559	( 32.1%)
G	356	( 20.5%)
C	508	( 29.1%)
Total	( 1740)	

(G-C)% = 61.2 %

\*\*\* GENIAS - Codon Usage Table \*\*\*  
 Frame No. 1  
 DNA Sequence Name (1): pEnza (1-1740)  
 Comment :

TTT Phe	4	0.7%	TCT Ser	1	0.2%	TAT Tyr	8	1.4%	TGT Cys	0	0.0%
TTC Phe	13	2.2%	TCC Ser	1	0.2%	TAC Tyr	14	2.4%	TGC Cys	0	1.0%
TTA Leu	0	0.0%	TCA Ser	0	0.0%	TAA stop	1	0.2%	TGA stop	0	0.0%
TTG Leu	4	0.7%	TGG Ser	23	4.0%	TAG stop	0	0.0%	TGG Trp	16	2.8%
CTT Leu	5	0.9%	CCT Pro	1	0.2%	CAT His	1	0.2%	CCT Arg	11	1.9%
CTC Leu	2	0.3%	CCC Pro	11	1.9%	CAC His	6	1.0%	CCG Arg	13	2.2%
CTA Leu	0	0.0%	CCA Pro	9	0.0%	CAG Gln	14	2.4%	CGA Arg	1	0.2%
CTG Leu	29	5.0%	CGG Pro	18	3.1%	CAG Gln	13	2.2%	CGG Arg	0	0.0%
ATT Ile	2	0.3%	ACT Thr	0	1.0%	AAT Asn	1	0.2%	AGT Ser	1	0.2%
ATC Ile	24	4.1%	ACC Thr	33	5.7%	AAC Asn	31	5.3%	AGC Ser	7	1.2%
ATA Ile	0	0.0%	ACA Thr	0	0.0%	AAA Lys	4	0.7%	AGA Arg	0	0.0%
ATG Met	18	2.8%	ACG Thr	12	2.1%	AAG Lys	3	0.5%	AGG Arg	0	0.0%
GTT Val	8	1.4%	GCT Ala	7	1.2%	GAT Asp	11	1.9%	GGT Gly	16	2.8%
GTC Val	26	4.5%	GCC Ala	23	4.0%	GAC Asp	24	4.1%	GCG Gly	44	7.6%
GTA Val	0	0.0%	GCA Ala	11	1.9%	GAA Glu	21	3.6%	GGA Gly	0	0.0%
GTG Val	12	2.1%	GCG Ala	9	1.6%	GAG Glu	12	2.1%	GGG Gly	0	0.0%
Total (360)											

Dc sm2 Ex3 1/2

Nov. 16 1988

A-1

4/1/70

570

52

### → Genetic Code [Universal]

A-2

42/70



A"-2

40/70

De Sm 2 Fx 1/2

250 280 270 280 290 300 310 320 330 340 350 360  
GTGACCCGCTTGTCCATGACGGCTCATATCTGGCAAAACCCCGGTGACGTGATCCAGGCCATCGAGCCCGGCAATCTGGAAACCCGCGCAATCGCGAAACATC  
V T P L V H D G V H Y L A N P G D V I Q A I D A A T G D L I V E H R Q L P H I  
370 380 390 400 410 420 430 440 450 460 470 480  
GCCACGCTGAACAGCTTTGGTGAGCGGACCGCGGCGATGCGCCCTATGACACCAACGCTTATTCGCTCTGGGCAACCACTTGGTTCGGCTGGACACCTCGACCGCGCAAGTCGTA  
A T L N S F G E P T R G M A L V G T N V Y F V S V D N H L V A L D T S T G Q V V  
490 500 510 520 530 540 550 560 570 580 590 600  
TTCCACGTCGATCCGGTCAAGGACGATGATGGTCTGAACTCTGTCGACCTTCGCGCCCGATTCGCGCAATGCGTCTGCGGCTCGACCTTCAGTATTCGCGCTTCGGCTGTTTCGTT  
F D V D R G Q G T D H V S N S G P I V A N G V I V A G S T C Q Y S P F G C F V  
610 620 630 640 650 660 670 680 690 700 710 720  
TCGGGGCCACGACTCGGCGCCGAGAGCTGTGGGCAACACCTTTATCCCGCGCGCGGGAAGGGTGATGAGACCTTGGGCAATATTACGAGCGCGCTGGATGACCGCGTT  
S G H D S A T G E L V R N (T, F, I, P, R, A, G, E, E, D, E, T, V, G, N, D, Y, E, A, R, V, H, T, G, V  
730 740 750 760 770 780 790 800 810 820 830 840  
TGGGGCCAGATCACCTATGACCCGTTGGCGGCTTGTCCACTACGACCTGACGAGTTGGCCCTCGGCGGAGATTACCGCGGCGACCGTTGGCGGCTCGATGTATGACCAACACC  
V G Q I T Y D P V G G L V H Y G T S A V G P A E I Q R G T V G G S H Y G T N T  
850 860 870 880 890 900 910 920 930 940 950 960  
CGCTTGTGTCGCGCCGAGAGATCGTCTGGCTGACCAACTTGCCTCGGCACTGGGACCAAGTGTACGTTGCGATGATGCTCAAGTTCAGCTCGACGTCAGCC  
R F A V R P E T G E I V V R H Q T L P R D N V D Q E C T P F E M N V V N V D V Q P  
970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080  
TCGGCTGAGTGAAGGCTGACGCCATCAACCCCGATCCCGGCGGAGGTCGCTTGTACCGCGCTTCCGTCGAAGACGACCGACCATGTTGGCAGTTTCGACCGCAACCGG  
S A E H E G L H A I N P D A A T G E R R V V T G V P C K N G T H V Q P D A E T G  
1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200  
GAATTCGTGGGCGCGACACGATATCAGAACCTGATGAAACGCTGATCCGATGCTGCTGATGTGAACGAAGATCTGGTGTGACCGAGCTGGAAGTGGCCTATGAAATC  
E F L V A R D T S Y Q H L I E S V D P D G L V H V N E D L V V T E L E V A Y E I  
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320  
TGCCGACCTTCTGGGTGGCGGCTGCGCTGCGCTGACCCCGATCTGGCATCTATTCATCCGCTGAACACGCTGTAGCGGTATGACGCTGTGACCAAGAGTTC  
C P T F L G G R D V P S A A L N P D T G I Y F I P L N N A C S G H T A V D Q E F  
1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440  
AGCTGCTGATGTGTATAACGTCACGCTGACTATAACTGTGCGCGGTTGGAACATGGCGCGTATCGACCGCATCGACATCGGCGCGCACCGCTGTGCTGCGCTGAACGC  
S S L D V Y N V S L D Y K L S P G S E N W G R I D A I D I S T G R T L V S A E R  
1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560  
TACGCTCGAACTACGCGCTGCTCTCCACCGCGGCGGCTGCTGTTCAACGGCGCACCGACCGTTACTTCGGGCGCCCTCAGGCAAGAGACCGCGGACGCTGTGGCAGACCCGT  
Y A S N Y A P V L S T G G G V L F N G G T D R Y F R A L S Q E T G E T L V Q T R  
1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680  
CTGCGGACTGTCGCTCGGTCAGCGATTTCCTATGATGACGCGGCTGCAATATGTCGCCATCGGCGCGCACCGATGCGGCAACCAACCGCGCCCTGACCGAGCGG  
L A T V A S G Q A I S Y E I D G V Q Y V A I G R G G T S Y G S N H N R A L T E R  
1690 1700 1710 1720 1730  
ATCGACTCGACCGCATCGCAGCGGCTATGCTTTGCTCTGCGCGGACGACTAAGCGACACAAAAGCGGTGAANAATCAGATCTCCGCGAGGTTCCGCGCTGCGGGGGGATTTT  
I D S T A I G S A I Y V F A L P Q Q  
TCATTGCGGGGCGCTCGACGGCGCTTAGGTTGGAGCCATCCGACGAGGCGCCAGATGTTTCCCATCGCATCGCCACATCCAGATGGGCGACATTCGCCATATTCACCGCGCTTGGGGG  
CAGTCTCGCCCCATCGCATGCCACCGCGCGCCCTTTGGCATCAGCGCGGACCGCTTGGG

## Sequences of the amplified products.

## 39F903 (697-1000)/A697f.Seq

TTNCGTGCCT GGGGCCAGAT CACCTATGAC CCCGTCACCA ACCTTGTCCA  
CTACGGCTCG ACCGCTGTGG GTCCGGCGTC GGAAACCCAA CGCGGCACCC  
CGGGCGGCAC GCTGTACGGC ACGAACACCC GTTTCGCCGT GCGTCCTGAC  
ACGGGCGAGA TTGTCTGGCG TCACCAGACC CTGCCCCGCG ACAACTGGGA  
CCAGGAATGC ACGTTCGAGA TGATGGTCAC CAATGTGGAT GTCCAACCCT  
CGACCGAGAT GGAAGGTCTG CAGTCGATCA ANCGAAANNN NNNNNNNNNN  
NNNNN

## 41F903 (697-1000)/A1000r.Seq

TTCCTCTTGG TCGAGGGTTG GACATCCACA TTGGTGACCA TCATCTCGAA  
CGTGCAATTCC TGGTCCCAGT TGTCGCGGGG CAGGGTCTGG TGACGCCAGA  
CAATCTCGCC CGTGTCAAGG CGCACGGCGA AACGGGTGTT CGTGCCGTAC  
AGCGTGCCGC CCGGGGTGCC GCGTTGGGTT TCCGACGCCG GACCCACAGC  
GGTCGAGCCG TAGTGGACAA GGTGGTGAC GGGGTCATAG GTGATCTGGC  
CCCAGGCACC GGTCATCCAA CGGGCTTTGT AANNNNNNNN NNNNNNNNNN  
N

## 43F903 (479-780)/A479f.Seq

AAAGCACTTT ATGGNCTCGA ACTCTCCGGC CCGATTGTCTG CCAATGGCGT  
CATCGTTGCG GGCTCGACCT GTCAGTATTC GCCGTTCCGC TGTTTCGTTT  
CGGGCCACGA CTCGGCCACC GGTGAAGAGC TGTGGCGCAA CACCTTTATC  
CCGCGCGCCG GCGAAGAGGG TGATGAGACC TGGGGCAATG ATTACGAGGC  
CCGCTGGATG ACCGGCGTTT GGGGCCAGAT CACCTATGAC CCCGTTGGCG  
GCCTTGTCCA CTACGGCACC TCAAGAGTTA ANANNNNNNN NNNNNNNNN

## 45F903 (479-780)/A780r.Seq

GACAAGGCTN NCACGGNGTC ATAGGTGATN TGGCCCCAAA CGCCGGTCAT  
CCAGCGGGCC TCGTAATCAT TGCCCCAGGT CTCATCACCC TCTTCGCCGG  
CGCGCGGGAT AAAGGTGTTG CGCCACAGCT CTTACCGGT GGCCGAGTCG  
TGGCCCCAAA CGAAACAGCC GAACGGCGAA TACTGACAGG TCGAGCCCGC  
AACGATGACG CCATTGGCGA CAATCGGGCC GGACGAGTTC GAGACCATAT  
CCGTGCCTTG ACCGCGATCG ACGTCCATAA ANNNNNNNNN NNNNNNNNN